

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-212505

(43)Date of publication of application : 15.08.1997

(51)Int.Cl.

G06F 17/27

(21)Application number : 08-034193

(71)Applicant : CANON INC

(22)Date of filing : 30.01.1996

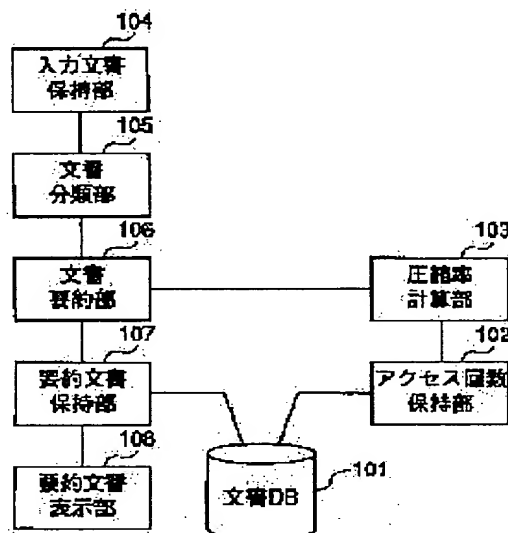
(72)Inventor : UEDA TAKANARI
OTANI NORIKO
ITO SHIRO
IKEDA YUJI

(54) DOCUMENT PROCESSOR AND ITS METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a document processor summarizing a document corresponding to the degree of user's interest.

SOLUTION: The document processor is provided with a document data base 101, the number of access holding part 102, a compressibility calculation part 103, an input document holding part 104, a document sorting part 105, a document summarizing part 106, a summarized document holding part 107 and a summarized document display part 108. A document sorting part 105 sort-processes an inputted document. The document sorting processing gives a sorting category to the document. The number of access to the given category is investigated and a compressibility calculation part 102 calculates a compressibility from the investigated number of access. The compressibility is made 10 when the number of access is not more than 10 times, the compressibility is made 3 when the number of access is not more than 100 times and the compressibility is made 2 when the number of access is not less than 101 times. The document summarizing part 106 summarizes the document according to the calculated compressibility. The summarized result is held in the summarized document holding part 107. The document held in the summarized document holding part 107 is displayed by the summarized document display part 108 by each sorting category.



LEGAL STATUS

[Date of request for examination] 27.01.2003

[Date of sending the examiner's decision of rejection] 01.03.2005

[Kind of final disposal of application other

than the examiner's decision of rejection or
application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's
decision of rejection]

[Date of requesting appeal against
examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-212505

(43) 公開日 平成9年(1997)8月15日

(51) Int.Cl.⁶

識別記号

庁内整理番号

F I

技術表示箇所

G 0 6 F 17/27

G 0 6 F 15/20

5 5 0 F

審査請求 未請求 請求項の数3 F D (全 6 頁)

(21) 出願番号 特願平8-34193

(22) 出願日 平成8年(1996)1月30日

(71) 出願人 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(72) 発明者 上田 隆也

東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内

(72) 発明者 大谷 紀子

東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内

(72) 発明者 伊藤 史朗

東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内

(74) 代理人 弁理士 渡部 敏彦

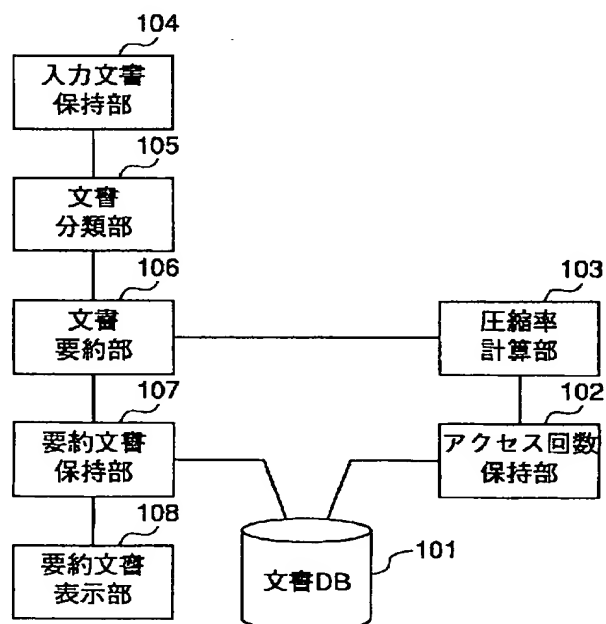
最終頁に続く

(54) 【発明の名称】 文書処理装置および方法

(57) 【要約】

【課題】 ユーザの興味の度合に応じて文書を要約する文書処理装置を提供する。

【解決手段】 文書処理装置は、文書データベース101、アクセス回数保持部102、圧縮率計算部103、入力文書保持部104、文書分類部105、文書要約部106、要約文書保持部107および要約文書表示部108を有する。文書分類部105は入力された文書についての文書分類処理を行う。文書分類処理は文書に分類カテゴリを付与する処理である。付与された分類カテゴリのアクセス回数を調べ、圧縮率計算部103は調べたアクセス回数から圧縮率を計算する。アクセス回数が10回までは圧縮率を10とし、アクセス回数が100回までは圧縮率を3とし、アクセス回数が101回以上である場合、圧縮率を2とする。文書要約部106は計算した圧縮率にしたがって文書を要約する。要約した結果は要約文書保持部107に保持される。要約文書保持部107に保持されている文書は分類カテゴリ毎に要約文書表示部108に表示される。



【特許請求の範囲】

【請求項 1】 入力された文書を保持する文書保持手段と、

該保持された文書を要約する要約手段とを備えた文書処理装置において、

前記文書保持手段に保持された文書に分類カテゴリを付与する分類付与手段と、

該付与された分類カテゴリ毎にユーザの興味の度合を判定する判定手段と、

該判定されたユーザの興味の度合に応じて圧縮率を設定する設定手段とを備え、

前記要約手段は該設定された圧縮率にしたがって前記文書を要約することを特徴とする文書処理装置。

【請求項 2】 前記判定手段は、前記文書が属する前記分類カテゴリにアクセスされるアクセス回数を計数する計数手段を備え、

前記設定手段は、該計数されたアクセス回数が多いほど前記圧縮率を低く設定することを特徴とする請求項 1 記載の文書処理装置。

【請求項 3】 入力された文書を保持し、
該保持された文書を要約する文書処理方法において、
前記保持された文書に分類カテゴリを付与し、
該付与された分類カテゴリ毎にユーザの興味の度合を判定し、
該判定されたユーザの興味の度合に応じて圧縮率を設定し、
該設定された圧縮率にしたがって前記文書を要約することを特徴とする文書処理方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は文書処理装置および方法に関する。

【0002】

【従来の技術】 近年、記憶媒体の大容量化・低価格化、ワードプロセッサの普及などによって電子化された文書の量が增大している。また、ネットワークの整備が進み、電子メール・電子ニュースなどのメディアによってユーザの下に届く電子化文書の量も増えている。このため、ユーザが処理できる量を越えた文書が入ってくるとい、いわゆる「情報洪水」が問題となっている。

【0003】 この問題の対応策として、各文書の量を減らすために文書の内容を要約する「文書要約」の技術が用いられるようになってきた。「文書要約」では、例えば文書中の各文に重要度にしたがってスコアを付け、スコアの高いものから文を選択することにより定められた比率で要約文書を生成する。ここではこの比率を圧縮率と呼ぶこととし、「原文書の量」と「要約文書の量」との比として定義する。

【0004】 即ち、圧縮率が高いほど要約文書の長さは短くなる。従来、圧縮率は装置側であらかじめ一定の値

に設定されていたり、あるいはユーザが直接に一定の値に設定するようにされていた。

【0005】

【発明が解決しようとする課題】 しかしながら、圧縮率は文書の種類によって違うことが望ましい。ユーザは興味の度合が高い種類の文書は詳細に内容を知りたい。一方、興味の度合がそれほどでもない種類の文書は簡潔に内容を知ることができればよい。一律の圧縮率が設定されている場合、こうした要求に応えることができない。

【0006】 また、ユーザが文書を見る際、圧縮率を任意の値に設定する手段が用意されている場合があるが、この場合、詳細に内容を知りたい種類の文書についてはいちいち圧縮率を低く設定して見なければならず、また、簡潔に内容を知りたい種類の文書についてはいちいち圧縮率を高く設定して見なければならない。

【0007】 そこで、本発明はユーザの興味の度合に応じて文書を要約する文書処理装置を提供することを目的とする。

【0008】

【課題を解決するための手段】 上記目的を達成するために、本発明の請求項 1 に係る文書処理装置は、入力された文書を保持する文書保持手段と、該保持された文書を要約する要約手段とを備えた文書処理装置において、前記文書保持手段に保持された文書に分類カテゴリを付与する分類付与手段と、該付与された分類カテゴリ毎にユーザの興味の度合を判定する判定手段と、該判定されたユーザの興味の度合に応じて圧縮率を設定する設定手段とを備え、前記要約手段は該設定された圧縮率にしたがって前記文書を要約することを特徴とする。

【0009】 請求項 2 に係る文書処理装置では、請求項 1 に係る文書処理装置において前記判定手段は、前記文書が属する前記分類カテゴリにアクセスされるアクセス回数を計数する計数手段を備え、前記設定手段は、該計数されたアクセス回数が多いほど前記圧縮率を低く設定することを特徴とする。

【0010】 請求項 3 に係る文書処理方法は、入力された文書を保持し、該保持された文書を要約する文書処理方法において、前記保持された文書に分類カテゴリを付与し、該付与された分類カテゴリ毎にユーザの興味の度合を判定し、該判定されたユーザの興味の度合に応じて圧縮率を設定し、該設定された圧縮率にしたがって前記文書を要約することを特徴とする。

【0011】

【発明の実施の形態】 本発明の文書処理装置の実施の形態について説明する。図 1 は実施の形態における文書処理装置の構成を示すブロック図である。図において、101 はユーザが保存した文書を格納する文書データベース、102 は文書データベース 101 の中の各分類カテゴリにアクセスした回数を保持するアクセス回数保持部である。

【0012】103は文書の圧縮率を計算する圧縮率計算部、104はユーザの下に入ってきた文書を保持する入力文書保持部、105は入力文書保持部104に保持されている文書を分類する文書分類部、106は圧縮率計算部103で計算した圧縮率にしたがって文書を要約する文書要約部である。

【0013】107は要約した文書を分類カテゴリ毎に保持する要約文書保持部、108は要約文書保持部107に保持されている文書を表示する要約文書表示部である。

【0014】図2は文書処理装置のハードウェアの構成を示すブロック図である。図において、201は後述する図3の制御手順を記憶する制御メモリである。制御メモリ201はROMあるいはRAMのいずれでもよい。202はメモリであり、前述のアクセス回数保持部102、入力文書保持部104、要約文書保持部107を実現する。

【0015】203は制御メモリ201に記憶されている制御手順にしたがって処理を行う中央処理装置である。204はディスクであり、文書データベース101を実現する。205はディスプレイであり、要約文書表示部108を実現する。ディスプレイ205はCRTであつてもよいし、液晶ディスプレイであつてもよい。206は各構成要素を接続するバスである。

【0016】図3は文書処理装置の動作の制御手順を示すフローチャートである。文書分類部105は入力された文書についての文書分類処理を行う（ステップS301）。文書分類処理は文書に分類カテゴリを付与する処理であり、一般に行われているどのような手法を用いてもよい。例えば、分類カテゴリ毎にキーワード集合を定めておき、入力された文書にどのようなキーワードが出現するかによって分類カテゴリを決定するようにしてもよい。

【0017】付与された分類カテゴリのアクセス回数を調べる（ステップS302）。アクセス回数は文書データベース101に保存されている各分類カテゴリの文書にアクセスした回数として定義できる。即ち、文書データベース101から文書を検索して利用する際、アクセス回数保持部102内に保持されているその文書が属する分類カテゴリのアクセス回数をインクリメントするようにしておく。このとき、アクセス回数の総数も同時にインクリメントする。

【0018】圧縮率計算部103はステップS302で調べたアクセス回数からステップS301で付与された分類カテゴリに対応する圧縮率を計算する（ステップS303）。アクセス回数の少ない分類カテゴリほど圧縮率が大きくなるように設定してあれば、どのような方式にしたがって圧縮率を設定してもよい。図4はアクセス回数に応じた圧縮率の値を示す説明図である。アクセス回数が10回までは圧縮率を10とし、アクセス回数が

100回までは圧縮率を3とし、アクセス回数が101回以上である場合、圧縮率を2とする。

【0019】文書要約部106はステップS303で計算した圧縮率にしたがって文書を要約する（ステップS304）。要約した結果は要約文書保持部107に保持される。

【0020】未処理の文書が残っているか否かを調べる（ステップS305）。未処理の文書が残っている場合、ステップS301の処理に戻って文書分類処理を行う。一方、全ての文書を処理し終わった場合、ステップS306に移行する。

【0021】ステップS306では、要約文書保持部107に保持されている文書を分類カテゴリ毎に要約文書表示部108に表示する。

【0022】図5は各分類カテゴリのアクセス回数を具体的に示す説明図である。図6は各文書の分類カテゴリおよびその圧縮率を示す説明図である。3つの文書がそれぞれ「Jリーグ」、「大相撲」、「プロ野球」に分類された場合、それぞれの圧縮率は図5のアクセス回数に基づき値2、10、3に設定される。即ち、文書1は分類カテゴリが「Jリーグ」であり、「Jリーグ」のアクセス回数は値107であるので、図4にしたがって圧縮率は値2となる。このようにして得られた圧縮率にしたがって各文書は要約される。

【0023】尚、上記実施の形態では、各分類カテゴリに対するユーザの興味の度合を分類カテゴリへのアクセス回数で捉えたが、その他の手法によってもよい。例えば、文書データベース中の各分類カテゴリに保存されている文書の数によってもよく、文書数が少ない分類カテゴリほど圧縮率を高く設定するようにする。

【0024】また、文書を表示する際、要約だけでなく選択によって全文も表示できるようになっている構成の場合、全文を表示させる回数の多い分類カテゴリはユーザの興味の度合が高いとみなすこともできる。全文を表示させる率、つまり、各分類カテゴリに属する文書のうち全文を表示させた文書の比率を保存しておき、その率が低いほど圧縮率を高く設定するようにする。

【0025】さらに、上記実施の形態では、アクセス回数から圧縮率を決定する際、実際にアクセスした回数の値を用いたが、全体のアクセス回数に対する比を用いてもよい。

【0026】また、上記実施の形態では、圧縮率はアクセス回数だけで決まるとしたが、文書の総数を考慮するようにしてもよい。例えば、その時点に表示する文書の総数を n 、アクセス回数から決まる圧縮率を r_0 、実際の圧縮率を r とすると数式1にしたがって設定してもよい。

【0027】

$$\begin{aligned} \text{【数1】 } r &= r_0 & : 0 < n \leq 50 \\ r &= r_0 \times 1.5 & : 50 < n \leq 100 \end{aligned}$$

$$r = r_0 \times 2 \quad : 100 < n$$

これにより、文書総数が多い場合、全体的に圧縮率を高めることができる。

【0028】さらに、上記実施の形態では、分類カテゴリ毎に圧縮率を設定して文書を要約するようにしていたが、要約文書の字数を設定してから圧縮率を算出するようにしてもよい。例えば、アクセス回数が101回以上の場合、400文字、アクセス回数が11回以上100回以下の場合、200文字、アクセス回数が10回以下の場合、100文字で要約するようにする。文字数 n で文字数 m の文書を要約する場合、その文書を圧縮率 m/n で要約すればよく、上記実施の形態と同じ文書要約の手法を使うことができる。

【0029】また、上記実施の形態では、圧縮率をアクセス回数から設定するようにしていたが、ユーザによって圧縮率を任意の値に設定できる手段が用意されている場合、各分類カテゴリに属する文書についてユーザが設定した圧縮率の履歴から圧縮率を設定するようにしてもよい。例えば、ユーザが分類カテゴリAに属する文書を見る際、ユーザが設定した圧縮率の平均を分類カテゴリAに対する圧縮率として使うことができる。

【0030】さらに、ユーザ自らが分類カテゴリ毎に興味の度合を数値で設定できるようにしてもよい。

【0031】また、文書総数が所定数以下である場合、文書の要約作成を中止するようにしてもよい。これにより、ユーザは文書の数が少ないときには即座に全文の内容を見ることができる。

【0032】さらに、本発明は複数の機器から構成されるシステムに適用してもよいし、1つの機器からなる装置に適用してもよい。また、本発明はシステムあるいは装置にプログラムを供給することによって達成される場合にも適用できることはいうまでもない。この場合、本発明を達成するためのソフトウェアによって表されるプログラムを格納した記憶媒体をシステムあるいは装置に読み出すことによってそのシステムあるいは装置が本発明の効果を楽しむことが可能となる。

【0033】

【発明の効果】本発明の請求項1に係る文書処理装置によれば、文書保持手段により入力された文書を保持し、要約手段により該保持された文書を要約する際、分類付与手段により前記文書保持手段に保持された文書に分類カテゴリを付与し、判定手段により該付与された分類カ

テゴリ毎にユーザの興味度合を判定し、設定手段により該判定されたユーザの興味度合に応じて圧縮率を設定し、前記要約手段は該設定された圧縮率にしたがって前記文書を要約するので、ユーザの興味度合に応じて文書を要約することができる。したがって、限られた時間内で効率的にユーザに情報を伝達できる。

【0034】請求項2に係る文書処理装置によれば、前記判定手段は、前記文書が属する前記分類カテゴリにアクセスされるアクセス回数を計数する計数手段を備え、前記設定手段は、該計数されたアクセス回数が多いほど前記圧縮率を低く設定するので、ユーザの興味度合を正しく判定することができる。

【0035】請求項3に係る文書処理方法によれば、入力された文書を保持し、該保持された文書を要約する文書処理方法において、前記保持された文書に分類カテゴリを付与し、該付与された分類カテゴリ毎にユーザの興味度合を判定し、該判定されたユーザの興味度合に応じて圧縮率を設定し、該設定された圧縮率にしたがって前記文書を要約するので、ユーザの興味度合に応じて文書を要約することができる。

【図面の簡単な説明】

【図1】実施の形態における文書処理装置の構成を示すブロック図である。

【図2】文書処理装置のハードウェアの構成を示すブロック図である。

【図3】文書処理装置の動作の制御手順を示すフローチャートである。

【図4】アクセス回数に応じた圧縮率の値を示す説明図である。

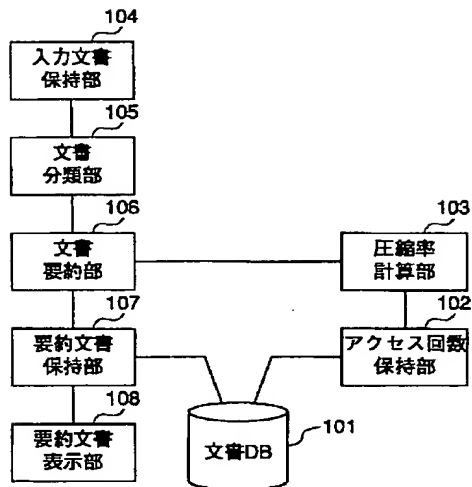
【図5】各分類カテゴリのアクセス回数を具体的に示す説明図である。

【図6】各文書の分類カテゴリおよびその圧縮率を示す説明図である。

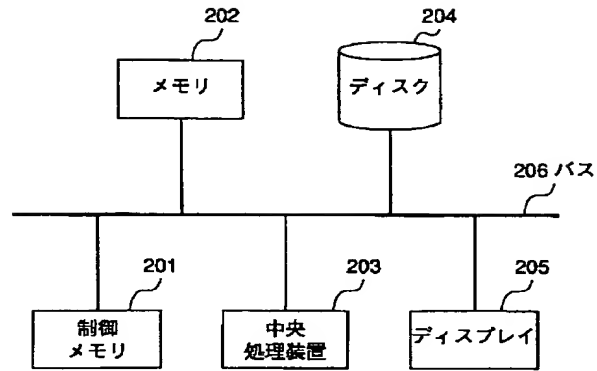
【符号の説明】

- 101 文書データベース
- 102 アクセス回数保持部
- 103 圧縮率計算部
- 104 入力文書保持部
- 105 文書分類部
- 106 文書要約部
- 107 要約文書保持部

【図 1】



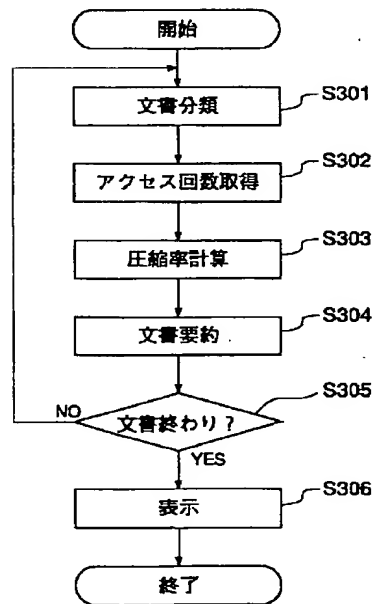
【図 2】



【図 5】

分類カテゴリ	アクセス回数
大相撲	4
Jリーグ	107
プロ野球	41
テニス	20

【図 3】



【図 4】

アクセス回数	圧縮率
～10回	10
～100回	3
101回以上	2

【図 6】

	分類カテゴリ	圧縮率
文書1	Jリーグ	2
文書2	大相撲	10
文書3	プロ野球	3

フロントページの続き

(72)発明者 池田 裕治

東京都大田区下丸子 3 丁目 30 番 2 号 キヤ
ノン株式会社内